Full Title: The Identification of a 1916 Irish Rebel: New Approach for Estimating Relatedness From Low Coverage Homozygous Genomes

Short Title: Estimating Relatedness From Low Coverage Homozygous Genomes

Daniel Fernandes[1,2*], Kendra Sirak[1,3], Mario Novak[1,4], John Finarelli[5,6], John Byrne[7], Edward Connolly[8], Jeanette EL Carlsson[5,9], Edmondo Ferretti[9], Ron Pinhasi[1], Jens Carlsson[5,9]

1 School of Archaeology, University College Dublin, Belfield, Dublin 4, Republic of Ireland

2 CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal

3 Department of Anthropology, Emory University, 201 Dowman Dr., Atlanta, GA 30322, United States of America

4 Institute for Anthropological Research, Ljudevita Gaja 32, 10000 Zagreb, Croatia

5 School of Biology and Environment Science, University College Dublin, Belfield, Dublin 4, Republic of Ireland

6 Earth Institute, University College Dublin, Belfield, Dublin 4, Republic of Ireland

7 National Forensic Coordination Office, Garda Technical Bureau, Garda Headquarters, Phoenix Park, Dublin 8, Republic of Ireland.

8 Forensic Science Ireland, Garda Headquarters, Phoenix Park, Dublin 8, Republic of Ireland

9 Area 52 Research Group, School of Biology and Environment Science, University College Dublin, Belfield, Dublin 4, Republic of Ireland

Corresponding author:

E-mail: dani.mag.fernandes@gmail.com (DF)

ABSTRACT

Thomas Kent was an Irish rebel who was executed by the British forces in the aftermath of the Rising armed insurrection of 1916 and buried in a shallow grave on Cork prison's grounds. In 2015, ninety-nine years after his death, a state funeral was offered to his living family to honor his role in the struggle for Irish independence. However, due to inaccuracies in record keeping, the bodily remains from Cork prison that supposedly belonged to Kent could not be identified with absolute certainty. At the request of the Irish National Police (An Garda Síochána), we identified the remains presumed to be those of Kent using a novel method that compared homozygous single nucleotide polymorphisms obtained from next-generation shotgun sequencing with genetic data from two of Kent's known living relatives who share with him a third degree relationship. DNA degradation resulted in low sequence depth of Thomas Kent's DNA and only a few loci appeared as heterozygous, rendering traditional methods of relatedness estimation unusable. To remove potential bias, all loci were therefore analysed as homozygous, leading to a process we refer to as "forced homozygote approach". This procedure reduces the expected degree of relatedness to half of what is expected when heterozygotes are included; thus, Kent had a relationship coefficient (Queller and Goodnight's Rxy) of 12.5% with his third-degree relatives instead of the expected 25%. The results were confirmed using simulated data for different relatedness classes (2000 simulated individuals per class), also confirming Thomas Kent's identity. We argue that this method provides a necessary alternative for relatedness estimations, not only in forensic analysis, but also in ancient DNA studies, where reduced amounts of genetic information can limit the application of traditional methods.

INTRODUCTION

Estimating the genetic relatedness of modern individuals is routinely achieved by employing the use of microsatellites (synonymous with short tandem repeats (STR)) or other genomic markers that estimate kinship coefficients based on probabilities of identity-by-descent (IBD) [1,2]. These methods, however, cannot be applied to cases where the DNA presents high levels of fragmentation and damage, as is common in ancient DNA (aDNA) research. Upon an organism's death, its genetic material starts to degrade and accumulate damage as repair enzymes no longer maintain the integrity of the molecular

structure of DNA [3,4]. Among the many factors that contribute to the rate and severity of this phenomenon are temperature, the acidity of the surrounding depositional environment, ambient level of humidity, and the eventual invasion of environmental microbes into the organism's cells. As a result, DNA fragments extracted from preserved tissue (in most cases bone and teeth) that is recovered from either ancient or semi-ancient (e.g. many forensic cases) human remains, are short in length, ranging from 30 to 70 base pairs. The degradation process has a major impact on the success rates and authenticity of many PCR-based ancient DNA (aDNA) identification techniques [3,4,5,6], however analysis of these short and damaged DNA molecules was revolutionised with the onset of Next Generation Sequencing (NGS) one decade ago. Next-Generation shotgun sequencing has enabled aDNA studies to progress at a much faster rate than before, and when applied in conjunction with optimised bone tissue isolation, DNA extraction, and sequencing technologies, large amounts of genetic information can be obtained even from samples with poor molecular preservation.

Relatedness estimation is a topic of relevance and interest in both anthropological and forensic studies. Before NGS, PCR-based studies were affected by a limited capacity to authenticate aDNA results and an inability to retrieve the required data from most aDNA samples [7,8,9,10]. However, some methods have been adapted to work specifically with this type of NGS or ancient DNA data; these are present in software such as PLINK2 [11] and NGSrelate [12]. Both software packages utilise Single Nucleotide Polymorphism (SNP) data, shown to work well with maximum likelihood approaches, and rely on genotypes, genotype likelihoods and minor allele frequencies. However, these packages require the input of relatively high amounts of genetic data (large numbers of loci) which is oftentimes challenging and expensive to obtain from ancient skeletal material [1,2,12]. Our method overcomes these challenges by substantially reducing the amount of input data required without sacrificing the confidence of the relatedness estimation. Here, we apply this novel method to identify the skeletal remains of a famous Irish Rebel, Thomas Kent.

Thomas Kent (1865-1916), an Irish rebel native to Castlelyons, grew up in Bawnard House located just outside the town of Fermoy in County Cork, Ireland. A week after the Easter Rising insurrection, in April of 1916, the Royal Irish Constabulary (RIC) raided the family home on 1st May. An RIC officer was shot dead during the raid. Thomas and William Kent were arrested. Following court martial, William was acquitted,

but Thomas received a death sentence and was one of 16 men executed by British Forces following the Easter Rising, being executed in the early hours of the 9th of May, 1916 at Cork Detention Barracks and then buried adjacent to where he fell [13].

The remains of Thomas Kent lay in the Barracks, which subsequently became Cork Prison, until June 2015, when they were exhumed by a team led by the National Monuments Service of the Department of Arts, Heritage and the Gaeltacht. Poorly kept records from the era of Thomas Kent's execution and throughout the intervening 99 years resulted in confusion surrounding his final resting place and uncertainty in the identification of his remains. The presumed identity of the remains was solely based on circumstantial evidence, and though attempted, it was soon determined that traditional DNA analysis was not an option due to the expected DNA degradation. The National Forensic Coordination Office at the Garda Technical Bureau and Forensic Science Ireland contacted the University College Dublin (UCD) who developed a new DNA identification method, based on optimisation techniques involving the use of the osseous inner ear part of the petrous part of the temporal bone [14] which has been applied successfully for over ~1000 archaeological samples from temperate regions spanning between 40,000-500 years before present (average endogenous yields range of 50-70% and with an overall success rate of ~80% [15]).

Using low-coverage shotgun sequencing data obtained from a single sequencing run on the Illumina MiSeq platform, we compared modern genetic data from two of Thomas Kent's living relatives to his century-old genetic material in order to identify his remains. Based on the success of the our analytic approach, we propose a NGS shotgun SNP-based method for relatedness estimation that uses "forced homozygote" allele data to estimate relationship coefficients and is based upon a symmetrical Rxy estimator algorithm developed by Queller and Goodnight [16].

Similar to other available softwares, the approach reported in this study relies on SNP data but requires a substantially lower amount of input data than the methods mentioned previously while not sacrificing any accuracy. This makes it widely applicable to the rapidly-expanding field of aDNA studies, where low coverage homozygous data is the norm, but also for time and budget-efficient forensic applications. Here we detail the success of our approach in the identification of the historical remains of the Irish revolutionary Thomas Kent.

RESULTS AND DISCUSSION

Authentication of Sequencing Data

As expected, DNA preservation differed noticeably between the modern individuals and the archaeological remains of Thomas Kent. The average sequence read length from Thomas Kent (TK) was predicted to be shorter than his modern relatives (E81 and E82) due to the historic nature of this sample; average fragment length was determined to be 54.01 base pairs (bp), with a wide standard deviation of ± 11.57bp (Table 1). In contrast, the modern relatives' DNA size averaged 64.48bp, with a standard deviation of ± 1.52bp. It should be noted that the average DNA fragment length is extremely close to the sequencing length used in the sequencing run (65bp). During the analysis of the raw sequencing data, the presence of adapters was detected in very few reads for the modern individuals as compared to the ancient sample (38% for E81 and E82, against 72% for TK), further supporting the notion that these endogenous DNA fragments were longer than 65bp. This was the expected outcome for modern DNA samples, indicating that these non-damaged sequences were possibly of lengths greater than or close to 65bp. Due to the archaeological nature of Thomas Kent's genetic material and the possibility of modern DNA contamination, raw data for this sample was first analysed to confirm the authenticity of the retrieved DNA as endogenous and ancient. To authenticate the DNA of TK as ancient, we utilised a widely-used approach developed for ancient DNA that quantifies deamination frequencies at the terminal ends of the DNA molecule, looking in particular for C>T substitutions at 5' overhangs that characterize the deamination of cytosines. Using the mapDamage v 2.0 software [17,18], the deamination frequencies present in TK's DNA, 0.14 C>T at the 5' end and 0.10 G>A at the 3' end (Figure 1) appear consistent with the expectation of molecular degradation for century-old bones interred in a shallow grave in the presence of quicklime (Barton 2010). In contrast, the modern DNA from TK's living relatives did not show significant damage patterns on the ends of the sequences. However, because fragments with shorter size than that of the sequencing length (65bp) are not expected to be overwhelmingly present in modern DNA, these deamination frequencies are not informative as it is probable that the ends of the molecules were not read.

Table 1: Sequencing data analysis and relatedness coefficient results

Individual	TK	E81	E82
Total reads	9168617	4817884	6081215
Trimmed reads	6603060	1805684	2335337
Aligned reads	2359538	3855705	4758208
Endogenous (%)	25.73	80.03	78.24
GC content (%)	37	40	40
Average bp size (stdv)	54.01 +- 11.57	64.48 +- 1.52	64.48 +- 1.45
MapDamage 5' | 3'	0.14 | 0.10	N/A | N/A	N/A | N/A
Molecular sex	Male	Female	Female
Coverage	0.04X	0.08X	0.1X
SNPs*[1]	17403	34195	42066
Common SNPs	E81-1328 / E82-1592	TK-1328 / E82-3840	TK-1592 / E81-3480
mt Haplogroup (score)	α1*[2] (0.77)	B5γ2*[2] (0.61)	B5γ2*[2] (0.61)
Relatedness coefficients	E81-0.1336 / E82-0.1236	TK-0.1336 / E82-0.2794	TK-0.11236 / E81-0.2794

*[1] From Haak's 2015 dataset

*[2] Real haplogroup information is not shown due to ethical constraints

Figure 1: DNA damage patterns from deamination frequencies of terminal bases

Mitochondrial haplogroups were also estimated for the three individuals (Table 1) using Phy-Mer [19]. For ethical reasons, the determined haplogroup is not reported in this paper; however, the two modern relatives, as expected, shared the same haplogroup, with a score of 0.61. This score estimates how well the given data matches the assigned haplogroup in the 0-1 interval. Thomas Kent's haplogroup was different from that of the relatives, with a score of 0.77. All three haplogroups were consistent with expectations for historic or modern individuals native to Ireland.

Relatedness Estimations

We estimated relatedness between the three members of the Kent family using very low coverage shotgun data (ranging from 0.04X to 0.1X) obtained from one MiSeq sequencing run, which currently generates a maximum of 25 million reads. Because we did not use a targeted enrichment or hybridization capture method to selectively identify and obtain common loci within the human genome, the output data for each individual was a random pool of overlapping reads. Along with the negative controls, these

three samples were the only samples placed on the sequencing run. Thomas Kent's DNA accounted for 46% of the run's total reads, with 2359538 of the 9168617 reads aligning to the human genome and representing a genomic coverage of 0.04X. Relative 1 had 4817884 total reads (24% of the run's total reads), with 3855705 aligning to the human genome (0.08X coverage) and Relative 2 slightly more, 6081215 total reads (30% of the run's total reads), from which 4758208 were of human origin (0.1X coverage) (Table 1). None of the negative controls rendered sequences. Using the dataset of SNPs developed for population and evolutionary genetic studies employed in [10], we called genotypes for 354,212 positions for each individual, obtaining 17403 SNPs called for Thomas Kent, 34195 for Relative 1 (E81), and 42066 for Relative 2 (E82). Out of these, we extracted only the shared SNPs between each dyad: Thomas Kent:Relative1 (1328 loci), Thomas Kent:Relative2 (1592 loci), and Relative1:Relative2 (3480 loci). As the total genome coverages were very low (Table 1), virtually all SNPs called had only one 1X read depth. Because we did not have more than one read per SNP position, we forced each SNP to be homozygous by repeating the called base to generate a diploid loci; this is referred to as the "forced homozygote" approach. For SNPs with more than 1X coverage, one call with phred quality above 30 was randomly selected and then "forced" homozygous by repeating the base as explained above. We estimated relationship coefficients for each of the three dyads using the Queller and Goodnight (1989) algorithm incorporated in the software SPAGeDi1-5a (build04-03-2015) [20], using the correspondent European allele frequencies downloaded from the 1000 Genomes website. As expected, the use of the forced homozygote approach resulted in relatedness coefficients (Rxy) of half the expected values. For the pair Relative1:Relative2, we observed an Rxy of 0.2794, consistent in modern genetics with third-degree relatedness, but with second-degree relatedness in the forced homozygote approach (i.e. equivalent to a modern genetics Rxy of 0.50). For Thomas Kent:Relative1, the Rxy was estimated at 0.1336, and for Thomas Kent:Relative2, it was estimated at 0.1236. These values are consistent with a third degree of relatedness for uncle/niece (25% in modern genetics, or 12.5% under our forced homozygote approach) between Thomas Kent and the two living relatives, supporting the positive identification of his remains.

The expected hypothesis that Thomas Kent was related to the two living relatives by a third-degree relationship and the two living relatives are related to each other by a second-degree relationship

(Hypothesis #4, Table 2) is unambiguously supported by the data, comprising nearly the entire posterior probability of the set of hypotheses. Using the posterior probabilities, the odds that this hypothesis is incorrect given the observed data is less than one in one million (8.15 E-07). Indeed, the Odds Ratio of the summed posterior probabilities for the four hypotheses proposing that the remains of Thomas Kent are related, in any manner, to both relatives versus the odds that he is unrelated to at least one of the two is in excess of 5 trillion, indicating conclusively that the remains found are related to the two living individuals.

Table 2: Set of potential relatedness hypotheses for the combinations of full sibling/parental and half sibling/uncle between the three subjects

Hypothesis #	TK - E81	TK - E82	E81 - E82	Model LnL	Posterior Probability
1	Unrelated	Unrelated	Unrelated	-99.3158	4.6096 E-48
2	Unrelated	Unrelated	Full Sibling	-5.58593	2.3444 E-07
3	Unrelated	Unrelated	Half Sibling	-32.6766	4.0244 E-19
4	Uncle	Uncle	Full Sibling	9.680128	0.9999
5	Parent	Parent	Full Sibling	-4.6796	5.8030 E-07
6	Parent	Uncle	Half Sibling	-23.3091	4.7092 E-15
7	Uncle	Parent	Half Sibling	-25.8717	3.6312 E-16
8	Parent	Unrelated	Unrelated	-97.7177	2.2789 E-47
9	Uncle	Unrelated	Unrelated	-91.8191	8.3072 E-45
10	Unrelated	Parent	Unrelated	-100.008	2.3079 E-48
11	Unrelated	Uncle	Unrelated	-91.5465	1.0910 E-44

CybRSex Package

In order to assess the accuracy of the relatedness estimations using forced homozygote data, we computed simulations on three relatedness classes - unrelated individuals, full siblings, and half siblings. We randomly generated SNP data for a total of 2000 virtual pairs of individuals using allele frequencies of the shared SNPs of each of the three possible dyads - Thomas Kent:Relative1, Thomas Kent:Relative2, Relative1:Relative2. For each of these combinations, 2000 unrelated individuals, 2000 full siblings, and 2000 half siblings were simulated, their relatedness coefficients calculated in SPAGeDI, and the

distribution visualised, as shown in Figure 2. The peaks of the curves are at the expected half-values of the relationship coefficients and it is clear that the results obtained for the three relative pairs fall within the expected ranges of variation.

Figure 2: Relatedness coefficients' distribution for virtual dyads.

"Forced homozygote" relatedness coefficients of computer generated individuals calculated using SPAGeDI1-5a, based on minor allele frequencies of the SNPs common to the pairs TK-E81, TK-E82, E81-E82. Orange-Unrelated, Green-Half Siblings, Blue-Full Siblings. Arrows indicate the halved "forced homozygote" relatedness coefficients found for each pair.

An R package (CybRSex) was developed to automate the two processes: simulations down to a true coefficient of relationship of 25%, and actual data tests. By using data in PLINK format as input, our package either runs SPAGeDI for the desired pairs of individuals, or generates X number of homozygous individuals for the given set of SNPs and their allele frequencies. It is then possible to visualise the distribution of simulated relatedness estimates for any given relatedness class with the expected ranges of variation from the specific input SNP data. The tests on pairs of individuals are performed by a function that requires two input files and an allele frequency file. Each of the three coefficients of relationship used for simulations (0%, 25% and 50%) have a pair of functions that generate the simulated individuals and then separately plots them as an histogram. One extra function retrieves the alleles used for a specific SPAGeDI test, and fits their frequencies to the required format for the virtual simulations. This package can be found at https://github.com/danimag/tkrelated.

CONCLUSIONS

A unique interdisciplinary research opportunity on this historical matter has allowed us to develop an efficient and accurate method for relatedness estimations using small amounts of genetic data. We were able to identify the skeletal remains of Thomas Kent, whose state funeral took place on the 18th of September of 2015, shortly after the identification of his remains. Applicable to both forensic and aDNA research, our method for relatedness estimations has important additional benefits when compared to

existing methods. When compared to the software packages PLINK and NGSrelate, the approach we present here requires substantially lower genomic coverage, which will prove helpful when large amounts of genomic data are unavailable. In a situation similar to ours, Korneliussen and Moltke [12], show that using the software NGSrelate to estimate relatedness based on a coverage of 1X results in large variance of relatedness estimates, yet it still performs better than PLINK. As we have shown, our method was effective with coverages ranging from 0.04X to 0.1X, an order of magnitude reduction on the amount of genetic data required. We also have designed an R package (CybRsex) to simulate virtual groups of (un)related individuals and their relatedness coefficients, based on Queller and Goodnights Rxy, from a given set of SNPs and corresponding allele frequencies. This should prove useful not only in routine forensic analysis but also in ancient DNA studies, where low endogenous DNA contents are often the norm. With the implementation of the "forced homozygote" method, estimating the relatedness between individuals in contexts such as multiple or mass burials may become a more routine task in future studies. This will benefit research in archaeology and anthropology, where the relationships of individuals found interred in multiple burials are often only hypothesized. In addition, we see broad applications of this new approach to forensic sciences, often tasked with identifying individuals interred in mass graves or with unknown but presumed identities.

MATERIALS AND METHODS

Archaeological Bone Sampling

To obtain genetic material from the skeletal remains of Thomas Kent, fine bone powder was retrieved from the cochlea of the left petrous part of the temporal bone that was detached from the rest of the cranium. While the petrous part of the temporal bone is accepted as yielding systematically higher endogenous DNA compared to other skeletal elements [21], the cochlea in particular was chosen because of research that demonstrated that the otic capsule, and particularly the cochlea, provides the highest endogenous DNA yield from any part of the petrous [14]. The powder was obtained using a minimally-destructive direct drilling technique developed at University College Dublin aimed at reducing any possible damage to the bone. A Dremel 9100 Fortiflex rotary tool, fitted with a small-sized spherical grinding bit (1.5mm) previously treated with bleach and ethanol, was set to medium speed and used to

obtain approximately 100mg of bone powder. The cochlea was accessed from the superior aspect of the petrous bone, limiting visible damage to a 2-3mm hole on the superior surface of the petrous. The bone powder generated from drilling the cochlear cavern was collected in a clean weighing boat and transferred to a 1.5mL sterile Eppendorf tube. This procedure was conducted in a clean sample preparation facility at UCD.

Blood Sampling and DNA Extraction for Modern Relatives

Blood samples were collected from Thomas Kent's living relatives in accordance with the prescribed methods employed by Forensic Science Ireland in the investigation of any unidentified remains. DNA extracts were then sent to University College Dublin for further processing. Informed consent was obtained by the Gardaí for the genetic analysis of this biological material.

DNA Extraction for Thomas Kent

DNA was extracted from Thomas Kent's bone powder following the protocol from [22] which improves upon the optimized silica-based extraction technique described in [6]. Extraction took place in a physically separated ancient DNA lab at UCD in adherence with stringent anti-contamination protocols. Approximately 50mg of bone powder was combined with 1mL of an extraction buffer solution containing 0.5M EDTA and Proteinase K (Roche Diagnostics). The bone powder was suspended by vortexing and incubated at 37°C with rotation for 18 hours in a ThermoMixer C (Eppendorf AG) and subsequently centrifuged for 2 minutes at 17.000 g in a Heraeus Pico 17 microcentrifuge (Thermo Scientific) to separate the undissolved bone from the supernatant solution. The supernatant solution was collected and added to 13mL of binding buffer solution containing guanidine hydrochloride (MW 95.53, 5M), isopropanol, Tween-20 (10%), and sodium acetate (3M) in a custom-made binding apparatus. This binding apparatus was constructed by forcibly fitting a reservoir removed from a Zymo-Spin V column (Zymo Research) into a MinElute silica spin column (Qiagen). This apparatus was then placed into a 50mL falcon tube [22]. The 14mL solution of binding buffer and DNA extract was added to the extension reservoir in the falcon tube, the cap was secured, and the falcon tube was centrifuged for 4 minutes at 2500rpm, rotated 90°, and centrifuged for another 2 minutes at 3,000rpm. The extension reservoir was then

disassembled and the MinElute column was placed into a 2mL collection tube. The column was dry-spun for 1 minute at 13,300rpm, and two wash steps were subsequently performed using 650µL of PE wash buffer. Finally, the column was placed into a clean 1.5mL Eppendorf tube and the DNA was eluted into 25µL of TET buffer.

DNA Library Preparation

Libraries for next-generation sequencing were built for all three DNA extracts using a modified version of [23] as outlined in [21], where blunt end repair was performed using NEBNext End-Repair (New England Biolabs Inc.) and Bst was inactivated by heat (20 minutes at 80°C). Thomas Kent's DNA library was prepared in a dedicated ancient DNA lab whereas the libraries for the DNA of two modern relatives were prepared in a modern DNA lab in UCD Earth Institute's Area 52. Indexing PCRs were performed with AccuPrime Pfx Supermix (Life Technology), with primer IS4 and an indexing primer. 3µL of the indexed library was added to 21µL of freshly prepared PCR mix, and combined with 1µL of unique index, enabling the pooling of samples for multiplex sequencing. This resulted in a final volume of 25µL. PCR amplification was performed using the following temperature cycling profile: 5 minutes at 95°C, 12 cycles of 15 sec at 95°C, 30 sec at 60°C, and 30 sec at 68°C, and a final period of 5 minutes at 68°C. PCR reactions were then purified using MinElute PCR Purification Kit (Qiagen), following the manufacturer's instructions. Assessment of the PCR reactions were performed on the Agilent 2100 Bioanalyzer following the guidelines of the manufacturer. Based on the concentrations indicated by the Bioanalyzer, samples were pooled in equimolar ratios for sequencing.

Next-Generation Sequencing

Libraries were sequenced on an Illumina MiSeq platform at the UCD Conway Institute of Biomolecular and Biomedical Research using 65 base pair (bp) single-end sequencing.

Bioinformatics Analysis

A custom ancient DNA bioinformatics pipeline written by the Pinhasi Lab was applied for processing of raw MiSeq data. The software cutadapt v1.5 [24] was used to trim adapter sequences. Minimum overlap

was set to 1 (- O 1) and minimum length to 17bp (-m 17). Alignment to the human reference genome (hg19, GRCh37) was processed by the Burrows-Wheeler Aligner [25] with disabled seed (-l 1000) and filtering for reads with a minimum quality QC score of 30. Duplicated sequences were removed using samtools v0.1.19-96b5f2294a [26]. To assess the authenticity of Thomas Kent's DNA as ancient, damage patterns were assessed using the mapDamage tool [18].

Single nucleotide polymorphisms were called using the Genome Analyzer Tool Kit's (GATK) Pileup tool for the 354,212 positions present in the Harvard's "Fully public genotype dataset" described in [10].

Relatedness Analysis

Most loci were represented by 1X reads, and this low read depth prevented identification of heterozygote loci for the vast majority of SNP loci in all three analysed individuals, although some loci had greater coverage. To be able to fully leverage the set of SNP loci, we modeled relatedness on a sample of single-read loci. For loci with greater read depth, we randomly selected one representative allele to reduce the bias that might have been introduced by allowing for some heterozygote loci. Thus, by ensuring that all loci contained only one allele we forced a "homozygote" structure on the data. We reduced the list of genotyped loci to only those loci shared for each dyad (i.e, Thomas Kent and Relative1, Thomas Kent and Relative2, Relative1 and 2). European allele frequencies at the shared loci for each comparison were retrieved from the 1000 genomes project (release 20100804 http://www.1000genomes.org/) using tabix (http://www.htslib.org/doc/tabix.html), and these were used as the reference frequencies for estimating degree of relatedness (symmetrical Rxy estimator, Queller and Goodnight 1989) using SPAGeDi1-5a (build04-03-2015) [20]. Forcing homozygosity on the data will necessarily impact the Queller & Goodnight coefficient, as only half the genome is being interrogated. This will reduce the anticipated relatedness between dyads by a factor of one-half (i.e, reducing full sibling/parent-offspring relationships from 0.5 to 0.25 and between niece or nephew-uncle or aunt/half sibling/grandchild-grandparent relationships from 0.25 to 0.125).

For the three dyads of relatedness comparison (Thomas Kent and Relative1, Thomas Kent and Relative2, Relative1 and 2), we simulated nine data sets, each with 2000 virtual pairs of full siblings, half siblings or

unrelated individuals, using the observed alleles held in common for the each of the comparisons and the correspondent European allele frequencies (release 20100804 http://www.1000genomes.org/). For each simulated data set, we forced the same homozygote condition, resulting in a comparable set of loci represented by one allele. We assessed the degree of relatedness for the simulated data sets with SPAGeDi1-5a. The output relatedness coefficient for each simulated data set was tabulated to create an empirical distribution for three degrees of relatedness (full sibling/parent, half sibling/uncle, unrelated) for the particular set of loci observed to be held in common for the three subjects.

The distribution of relatedness coefficients was nearly normal (Figure 2). Using mean and variance parameters fit to the empirical distributions, we calculated maximum likelihood (ML) fits of the observed degree of relatedness for each dyad to the three relatedness distributions. Potential relatedness hypotheses constitute the set of potential combinations of full sibling/parental and half sibling/uncle between the three subjects, producing a set of eleven potential hypotheses (Table 2). The ML fit of each hypothesis is then the sum of the ML fits of the observed relatedness coefficient between the two individuals for the appropriate empirical distribution.

Ethics Statement

The investigation into the authentication of Thomas Kent’s remains was tasked to the Irish Police, An Garda Siochana, on behalf of the State, and therefore obeyed to specific ethical and legal considerations.

Informed consent was obtained when collecting the blood from Thomas Kent's living relatives in regard to analysis of the genetic data and dissemination of the results. Kent's remains, legally considered archaeological, were handled with the permission of the correspondent legal authorities

The request for assistance from UCD, by An Garda Siochana, to identify the remains recovered from Cork Prison in early 2015 was made to progress that element of the overall investigation.

An Garda Siochana are tasked with such investigations, on behalf of the State, and do not require an ethics committee to initiate enquiries. An Garda Siochana may enlist the expertise of any agency or academic entity to pursue lines of enquiry and such was the case with UCD.

In this case, the original request for help in identifying the remains came from the Department of An Taoiseach (Head of Government) to the National Forensic Coordination Office, who then managed the overall investigation. The integrity of all evidence, samples and results was managed by the Head of the National Forensic Coordination Office, who was also the investigating officer in this case. All ethical considerations and legal obligations under the Data Protection Acts were his responsibility as Investigating Officer, and he was the person who sought the assistance of UCD on behalf of An Garda Siochana. He reviewed all evidence or results before they were communicated to the relevant parties. In such investigations ethical considerations form part of the overall review, in addition to many layers of legal consideration and all requirements were met.

ACKNOWLEDGEMENTS

We would like to thank Dr. Sudipto Das for his comments on sequencing methods; the Irish Government for their support throughout the Thomas Kent identification process; and Dr. Olivia Cheronet for helping with the "CybRsex" R package.

REFERENCES

1. Powell JE, Visscher PM, Goddard ME. Reconciling the analysis of IBD and IBS in complex trait studies. Nat Rev Genet. 2010;11(11):800-5.

2. Speed D, Balding DJ. Relatedness in the post-genomic era: is it still useful? Nat Rev Genet. 2015;16(1):33-44.

3. Pääbo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences. 1989;86(6):1939-43.

4. Mitchell D, Willerslev E, Hansen A. Damage and repair of ancient DNA. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2005;571(1-2):265-76.

5. Willerslev E, Hansen AJ, Rnn R, Brand TB, Barnes I, Wiuf C, et al. Long-term persistence of bacterial DNA. Current Biology. 2004;14(1):R9-R10.

394 6. Rohland N, Hofreiter M. Ancient DNA extraction from bones and teeth. Nat Protocols.
395 2007;2(7):1756-62.

396 7. Baca M, Doan K, Sobczyk M, Stankovic A, Węgleński P. Ancient DNA reveals kinship burial patterns
397 of a pre-Columbian Andean community. BMC Genetics. 2012;13(1):1-11.

398 8. Deguilloux MF, Pemonge MH, Mendisco F, Thibon D, Cartron I, Castex D. Ancient DNA and kinship
399 analysis of human remains deposited in Merovingian necropolis sarcophagi (Jau Dignac et Loirac,
400 France, 7th-8th century AD). Journal of Archaeological Science. 2014;41:399-405.

401 9. Dudar JC, Waye JS, Saunders SR. Determination of a kinship system using ancient DNA, mortuary
402 practice, and historic records in an upper Canadian pioneer cemetery. International Journal of
403 Osteoarchaeology. 2003;13(4):232-46.

404 10. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the
405 steppe was a source for Indo-European languages in Europe. Nature. 2015;522(7555):207-11.

406 11. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to
407 the challenge of larger and richer datasets. GigaScience. 2015;4(1):1-16.

408 12. Korneliussen TS, Moltke I. NgsRelate: a software tool for estimating pairwise relatedness from
409 next-generation sequencing data. Bioinformatics. 2015;31(24):4009-11.

410 13. Barton B. The Secret Court Martial Records of the Easter Rising. Dublin: The History Press Ireland;
411 2010.

412 14. Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, et al. Optimal Ancient
413 DNA Yields from the Inner Ear Part of the Human Petrous Bone. PloS one. 2015;10(6):e0129102.

414 15. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide
415 patterns of selection in 230 ancient Eurasians. Nature. 2015;528(7583):499-503.

416 16. Queller DC, Goodnight KF. Estimating relatedness using genetic markers. Evolution.
417 1989;43:258-75.

418 17. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: testing for damage
419 patterns in ancient DNA sequences. Bioinformatics. 2011;27(15):2153-5.

420 18. Jnsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate
421 Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013;29(13):1682-4.

422 19. Navarro-Gomez D, Leipzig J, Shen L, Lott M, Stassen APM, Wallace DC, et al. Phy-Mer: a novel
423 alignment-free and reference-independent mitochondrial haplogroup classifier. Bioinformatics.
424 2015;31(8):1310-2.

425 20. Hardy OJ, Vekemans X. spagedi: a versatile computer program to analyse spatial genetic structure
426 at the individual or population levels. Molecular Ecology Notes. 2002;2(4):618-20.

427 21. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome
428 flux and stasis in a five millennium transect of European prehistory. Nature communications.
429 2014;5:5257.

430 22. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, et al. Complete mitochondrial
431 genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments.
432 Proceedings of the National Academy of Sciences. 2013;110(39):15758-63.

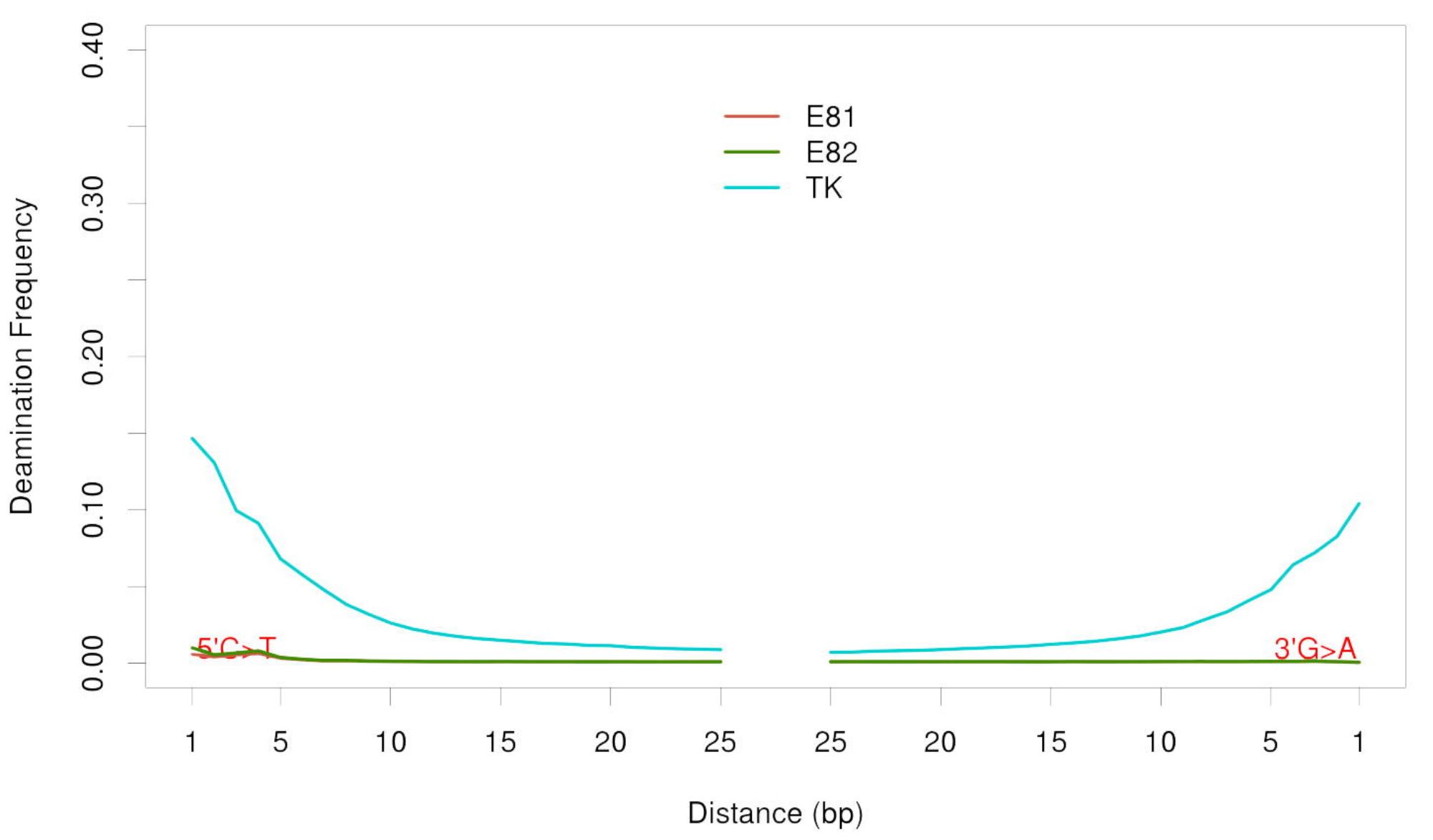
433 23. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture
434 and sequencing. Cold Spring Harbor protocols. 2010;2010(6):pdb prot5448.

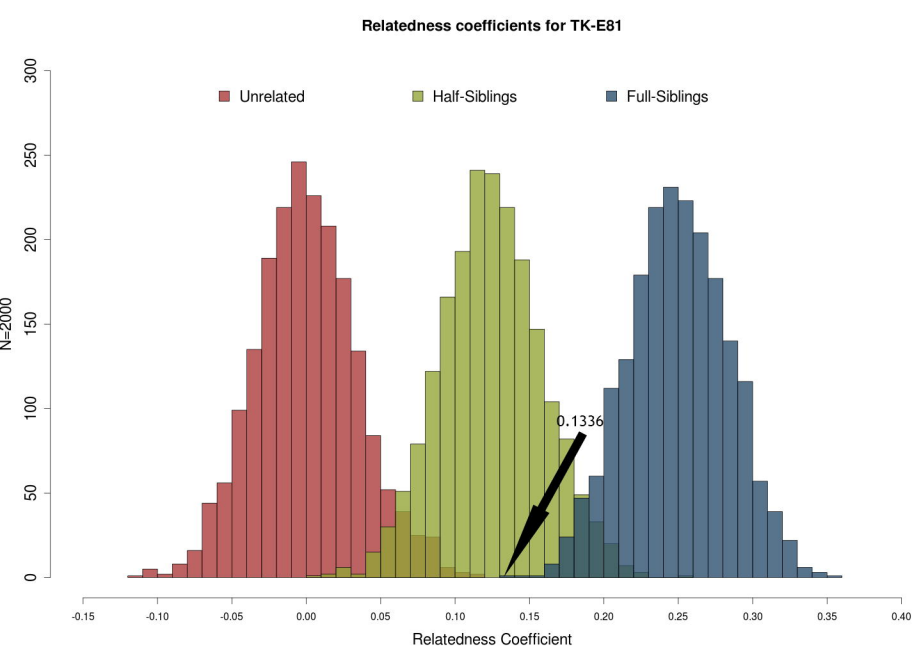
435 24. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011.
436 2011;17(1).

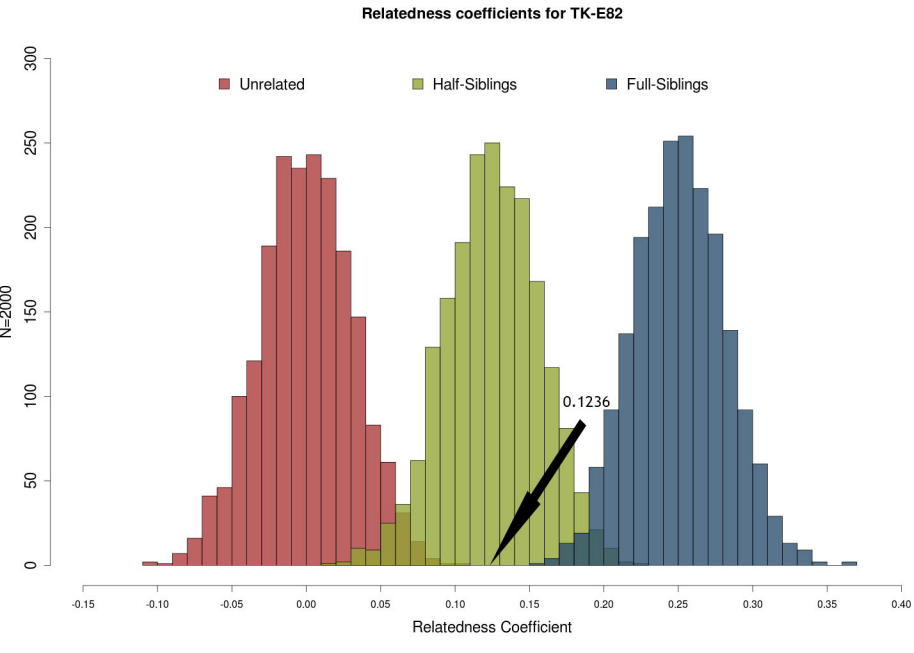
437 25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
438 Bioinformatics. 2009;25(14):1754-60.

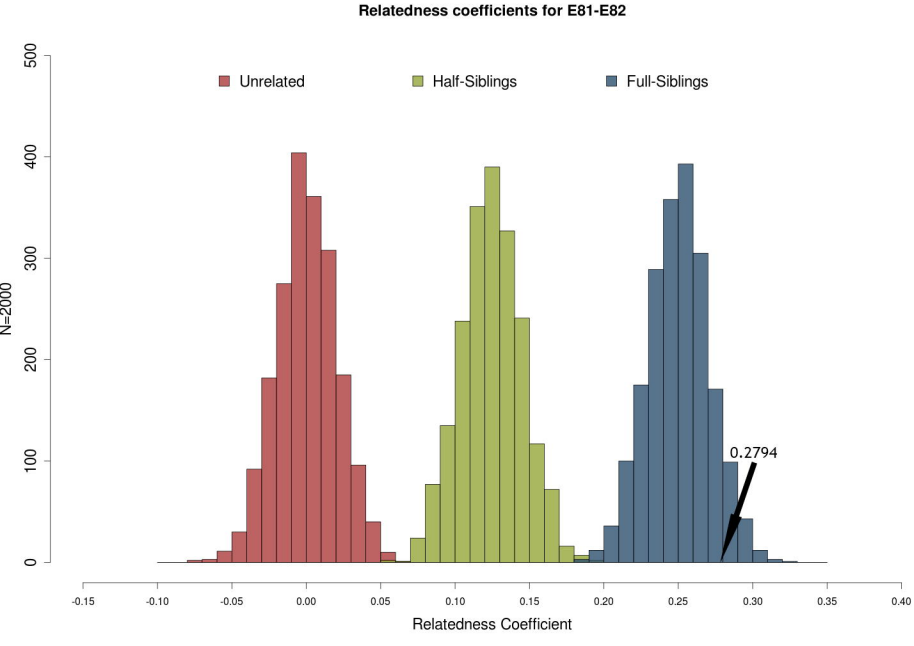
439 26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
440 format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

441


Deamination Frequency
0.00
0.10
0.20
0.30
0.40
E81
E82
TK
5'C>T
3'G>A
1
5
10
15
20
25
25
20
15
10
5
1
Distance (bp)


Relatedness coefficients for TK-E81
Unrelated
Half-Siblings
Full-Siblings
N=2000
0.1336
Relatedness Coefficient


Relatedness coefficients for TK-E82
Unrelated
Half-Siblings
Full-Siblings
N=2000
0.1236
Relatedness Coefficient


Relatedness coefficients for E81-E82
Unrelated
Half-Siblings
Full-Siblings
N=2000
0.2794
Relatedness Coefficient